

| Result No. | Score | Query |        | ID         | Description        |
|------------|-------|-------|--------|------------|--------------------|
|            |       | Match | Length |            |                    |
| 1          | 133   | 5.7   | 223    | FARE_METYA | Q68896 methanococc |
| 2          | 120.5 | 5.1   | 232    | FARE_METTH | Q26350 methanobact |
| 3          | 111.5 | 4.7   | 695    | FSHR_WACFA | P32212 macaca fasp |
| 4          | 109   | 4.6   | 693    | FSHR_CHICK | P79763 gallus gall |
| 5          | 109   | 4.6   | 712    | FRB6_YEAST | Q12473 saccharomyc |
| 6          | 107.5 | 4.6   | 629    | FRF7_YEAST | Q12343 saccharomyc |
| 7          | 107.5 | 4.6   | 695    | FSHR_HUMAN | P23945 homo sapien |
| 8          | 104.5 | 4.4   | 695    | FSHR_PIG   | P49059 sus scrofa  |
| 9          | 104   | 4.4   | 396    | YDEA_SCOLI | P31122 escherichia |
| 10         | 103   | 4.4   | 465    | NPT1_MOUSE | Q61983 mus musculu |
| 11         | 100   | 4.3   | 476    | RF3_SACBA  | P05512 saccharomyc |
| 12         | 99.5  | 4.2   | 452    | NUAM_BRALA | O79421 branchiosto |
| 13         | 99    | 4.2   | 365    | GP68_HUMAN | Q15743 homo sapien |
| 14         | 98    | 4.2   | 3411   | POIG_YEFV1 | P03314 y genome po |
| 15         | 98    | 4.2   | 3411   | POIG_YEFV2 | P19901 y genome po |
| 16         | 97.5  | 4.1   | 695    | FSHR_BOVIN | P33376 bos taurus  |
| 17         | 97    | 4.1   | 692    | FSHR_RAT   | P20395 rattus norv |
| 18         | 96.5  | 4.1   | 686    | FRSL_YEAST | P32791 saccharomyc |
| 19         | 96    | 4.1   | 261    | PHSC_ECOLI | P77409 escherichia |
| 20         | 95    | 4.0   | 393    | NUCC_ARATH | P56753 arabidopsis |
| 21         | 94.5  | 4.0   | 492    | SECY_CYAPA | P25014 cyanophora  |
| 22         | 94.5  | 4.0   | 695    | FSHR_SHEEP | P35379 oris aries  |
| 23         | 93.5  | 4.0   | 452    | NUAM_BRAEL | O47423 branchiosto |
| 24         | 93.5  | 4.0   | 499    | GSHR_PLAF7 | OL5770 plasmodium  |
| 25         | 93.5  | 4.0   | 1233   | VLC1_REOVD | HL5770 plasmodium  |
| 26         | 93    | 4.0   | 388    | HMC3_DESVH | P33390 desulfovibr |
| 27         | 93    | 4.0   | 459    | NUAM_BOVIN | P30910 bos taurus  |
| 28         | 93    | 4.0   | 521    | YT25_CAEEL | Q10934 caenorhabdi |
| 29         | 92.5  | 3.9   | 490    | GIFL_CHICK | P46896 gallus gall |
| 30         | 92.5  | 3.9   | 883    | YHL6_YEAST | P37811 saccharomyc |
| 31         | 92.5  | 3.9   | 1769   | YUK9_YEAST | P42945 saccharomyc |
| 32         | 92.5  | 3.9   | 2037   | FAS1_CANAL | P34731 c fatty aci |
| 33         | 91.5  | 3.9   | 495    | TPS1_YEAST | Q00764 s alpha.alp |





FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 4 4 G -> D (IN REF. 2).  
 FT CONFLICT 88 88 G -> A (IN REF. 2).  
 FT CONFLICT 140 140 K -> R (IN REF. 2).  
 FT CONFLICT 174 174 I -> T (IN REF. 2).  
 FT CONFLICT 191 191 N -> S (IN REF. 2).  
 FT CONFLICT 329 329 V -> L (IN REF. 2).  
 SQ SEQUENCE 693 AA; 78697 MW; 45F98699635A1BEC CRC64;

Query Match 4.68; Score 109; DB 1; Length 693;  
 Best Local Similarity 17.28; Pred. No. 1.3;  
 Matches 100; Conservative 72; Mismatches 173; Indels 238; Gaps 21;

QY 18 LPNG-INGIDARKVTVG-----VIGSGDPAKSLTIRLCGYHVVIGSRNPKFASFPF 71  
 DB 61 IPKCAFTGLHDLKIEISONDALIEIGNVFSSL-----PKL----- 97  
 QY 72 HVDVTHEDALTNIIFVAIHREHYTSLWDLRLHVG-----KILI 114  
 DB 98 -----HEIRIEKANKL-MKIDODAFQHLPLSLRYLLISNTGLSFLPVVHKVHSFQKVL 149  
 QY 115 DYSNNMINEYPSNAEYASLPDPSLIIVKGFNVSAWALQGLPKDASQVYICSNIIQA 174  
 DB 150 DVODNIHIRTIERNTFGLSS-----ESVILR-----LNKNGIQE 184  
 QY 175 ROOVIELAROLFIPIDGLSSAREIENLPLRLFTLWRGPVVAISLATFFFLYSFVRD 234  
 DB 185 IKD-----HAFNGTCLDELNLSDNYNLEKLPKVFQGAIGPVVLDISRTISFLPSHGLE 239  
 QY 235 VIHPYARNOOSDFKIPFIEIVNKTLPVATLGLSVLAGL----- 275  
 DB 240 FIKKLARSTYKALKLP--DVNKRSLIEANTYPSHCCAFNTRKNTONTIEFYPICSMSPA 297  
 QY 276 -----LAAAYLYGKY-----RRF-----PPWLETWLQCR 302  
 DB 298 KODLGEOTGRKRRSAAEDYISHYTRFGVNEFQYDGLCNEVDVDCSPKDPANPCE 357  
 QY 303 KQLG-----LLSFPFAMHVA-----YSLCLPMRRSERVFLNMAVQVQVHAN 344  
 DB 358 DIMGYNLVRLINILAITNGTNTVLLILISSQKLTVP-----RELMLNLAFAO----- 408  
 QY 345 IENSWNEEYWRIDWYISFGIMSLGLSLAVTSPVSNALNWR-----FSFIQ 395  
 DB 409 -----LCIGIYLFLFIASVQIDQTSRYNYAIDQOTGAGCNAAGFFTVFA 452  
 QY 396 STLGYVALLSTPH-----VLYGWKEAF----- 419  
 DB 453 SELSVYTLVTITLERWHITTYAMOLNKKVRLNMAVIMVFGMFAFTVALLPIFGTSSYM 512  
 QY 420 -----EEYRYEYTPNFVLAFLVLPISVILDLQLCRY 452  
 DB 513 KVSICLPHIETPSQAYV--IFLLVNLVLAFLVLCIYICICIY 553

## RESULT 5

FR6\_YEAST  
 ID FR6\_YEAST STANDARD; PRT; 712 AA.  
 AC Q12473;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE FERRIC REDUCTASE TRANSMEMBRANE COMPONENT 6 PRECURSOR.  
 GN FR6 OR YLL051C OR L0593.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
 RC Saccharomycetaceae; Saccharomycetes.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RA Wedler H., Wambutt R.;  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
 CC -|- COFACTOR: FAD (PROBABLE).

CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -|- SIMILARITY: BELONGS TO THE FRE / CYBB FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z47973; CAA88006.1; -  
 CC EMBL; Z73156; CAA97503.1; -  
 CC SGD; S0003974; FREG.  
 CC INTERPRO: IPR002916; -  
 CC PFAM; PF01794; Ferric\_reduct; 1.  
 KW Electron transport; Transmembrane; Iron transport; FAD; HAD;  
 KW Glycoprotein; Signal; Multigene family.  
 FT SIGNAL 1 17 POTENTIAL.  
 FT CHAIN 18 712 FERRIC REDUCTASE TRANSMEMBRANE COMPONENT  
 FT 6.  
 FT NP\_BIND 493 499 FAD (POTENTIAL).  
 FT TRANSMEM 168 188 POTENTIAL.  
 FT TRANSMEM 245 265 POTENTIAL.  
 FT TRANSMEM 288 308 POTENTIAL.  
 FT TRANSMEM 329 349 POTENTIAL.  
 FT TRANSMEM 361 381 POTENTIAL.  
 FT TRANSMEM 388 408 POTENTIAL.  
 FT TRANSMEM 417 437 POTENTIAL.  
 FT TRANSMEM 478 498 POTENTIAL.  
 FT TRANSMEM 550 570 POTENTIAL.  
 FT DOMAIN 553 556 POLY-LEU.  
 FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 450 450 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 655 655 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 712 AA; 81989 MW; 5224F12B51544BAA CRC64;

Query Match 4.68; Score 109; DB 1; Length 712;  
 Best Local Similarity 18.88; Pred. No. 1.3;  
 Matches 72; Conservative 51; Mismatches 119; Indels 142; Gaps 17;

QY 88 IIRVAIHREYTSL-----WDLRLHVGKILIDVSNMNRINQYSPESNAEYASL 136  
 DB 183 IIAVFFHMSHYNGLNRLALPASRFVNYIRGFVLPFLVD-----KHANFRFLNVEFTGL 238  
 QY 137 FPDLSIVKGFNVSAWALQGLPKDASQVYICSNIIQARQOVIELAROLFIPIDGLSLS 196  
 DB 239 MPNSL-----EAW----- 246  
 QY 197 SAREIENLPLRLFTLWRGPVVAISLATFFFLYFVRDVIHPYARNOOSDFKIPFIEIV 255  
 DB 247 -----IFGYTLANILFSLISYIIDPYNLIFNLSQFTRL-----LA 284  
 QY 256 NKTLPVATLTLVLVLAGLAAAYLYGKYKRRPPWLETWLQCRKQLGLSFFFAW 315  
 DB 285 DRS-GILAPTOFPLIIIFTARNSEFLFTGVKNSF-----ISFKWGRIMVLNATI 336  
 QY 316 H-VAYSICLPMRRSERVFLNMAVQVQVHANTENSWNEEYWRIDWYISFGIMSLGLSLL 374  
 DB 337 HSLSYSL-----FALINHAFK-----ISNK-----QLYWKFGIASITVLCVL 373  
 QY 375 AVTSIPSVSNALNWRSEFTOSTLGYVALLI---STPHVLIY-GWKRAFE-----EE 422  
 DB 374 LVLSLGIIVRK-----RHVEFFLYTHILALLFFYCCMOHVKIFNGKWEIVYVSLINGLEK 429  
 QY 423 YRFRYTPNFVLAFLVLPISVILDL 446  
 DB 430 LFRINW-----ILOFRFPKATLINL 449

```

RESULT 6
FRET_YEAST
ID FRET_YEAST STANDARD PRT 629 AA.
AC Q12333;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE FERRIC REDUCTASE TRANSMEMBRANE COMPONENT 7 PRECURSOR.
GN FRET OR YOL152W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=96132030; PubMed=8553699;
RA Casamayor A., Aldea M., Casas C., Herrero E., Gamo F.J.,
RA Lafuente M.J., Gancedo C., Arino J.
RT "DNA sequence analysis of a 13 kbp fragment of the left arm of yeast
RT chromosome XV containing seven new open reading frames."
RL Yeast 11:1281-1288(1995).
CC -!- COFACTOR: FAD (PROBABLE).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE FRET / CYBB FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; Z48239; CAA8276.1; -
DR EMBL; Z74894; CAA9174.1; -
DR SGD; S0005512; FRET;
DR INTERPRO; IPR002916; -
DR PFAM; PF01794; Ferric_reduct; 1.
DR KEGG; Electron transport; Transmembrane; Iron transport; FAD; NAD;
DR KEGG; Glycoprotein; Signal; Multigene family.
FT SIGNAL 1 ?
FT CHAIN 1 ? 629
FT FT FERRIC REDUCTASE TRANSMEMBRANE COMPONENT
FT FT 7.
FT NP_BIND 369 375 FAD (POTENTIAL).
FT TRANSMEM 46 66 POTENTIAL.
FT TRANSMEM 108 128 POTENTIAL.
FT TRANSMEM 168 188 POTENTIAL.
FT TRANSMEM 195 215 POTENTIAL.
FT TRANSMEM 238 258 POTENTIAL.
FT TRANSMEM 266 286 POTENTIAL.
FT TRANSMEM 293 313 POTENTIAL.
FT TRANSMEM 422 442 POTENTIAL.
FT CARBOHYD 330 330 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 541 541 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 629 AA; 71996 MW; 2384480E289C16F CRC64;

Query Match 4.6%; Score 107.5; DB 1; Length 629;
Best Local Similarity 21.7%; Pred. No. 1.4;
Matches 57; Conservative 46; Mismatches 77; Indels 83; Gaps 17;

Qy 214 GPVVVAISLATEFFLYSVRDVVIHYPYARNOQSDPKYKIPVINKVLPVAVITLISLVY-L 272
Db 123 GPFLVMATILYLYCFVP---HPFYR-PCAGFGSPPLSV---RAGIMASLSPVFEVSL 175
Qy 273 AG-----LLAAAYQLYGGKGYRFPWLETLWOCRLGLLSLFFANVHVAYSCLPLM 325
Db 176 SKKNIGWGLVLSYE-----KINHYH-----ASLCLFFSWVHV-----IPF 215
Qy 326 RRSRYFLNNAYQQVHANIEENWEEVRIEMYSIFGIMSLGLSLAVTSIP----- 380
Db 216 LQARH---EGGYERN---QRWKASDWR-----SGVPPFLNLLWLSLPIARRH 262

```

```

Qy 381 --SVSNALNREFSEFIQSTLGYVALLISTFHV-----LIYWKRAFEEYR- 425
Db 263 FYEIFLQLHW-----ILAVGYISLF---YHYPELNSHMYLVATVWV---FAQLFVRL 311
Qy 426 ---PYTPPNEFLALVLPISVIL 444
Db 312 AVKGYLRGPRFSFMASITIANVSIV 334

RESULT 7
FSHR_HUMAN
ID FSHR_HUMAN STANDARD PRT 695 AA.
AC P23945;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE FOLLICLE STIMULATING HORMONE RECEPTOR PRECURSOR (FSH-R) (FOLLITROPIN
DE RECEPTOR).
GN FSHR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=OVARY;
RX MEDLINE=91222171; PubMed=1709010;
RA Minegishi T., Nakamura K., Takakura Y., Ibukki Y., Igarashi M.;
RT "Cloning and sequencing of human FSH receptor cDNA."
RL Biochem. Biophys. Res. Commun. 175:1125-1130(1991).
RN [2]
SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=93245012; PubMed=1301382;
RA Kelton C.A., Cheng S.V., Nugent N.P., Schweickhardt R.L.,
RA Rosenthal J.L., Overton S.A., Wands G.D., Kuzaja J.B., Luchette C.A.,
RA Chappel S.C.;
RT "The cloning of the human follicle stimulating hormone receptor and
RT its expression in COS-7, CHO, and Y-1 cells."
RL Mol. Cell. Endocrinol. 89:141-151(1992).
RN [3]
SEQUENCE FROM N.A.
RA Tilly L.T., Alhara T., Nishimori K., Jai X.-C., Billig H.,
RA Kowalski K.I., Perlas E.A., Hsueh A.J.;
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
SEQUENCE OF 1-342 FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=93075197; PubMed=1359889;
RA Gromoll J., Gudermann T., Nieschlag E.;
RT "Molecular cloning of a truncated isoform of the human follicle
RT stimulating hormone receptor."
RL Biochem. Biophys. Res. Commun. 188:1077-1083(1992).
RN [5]
SEQUENCE OF 1-51 FROM N.A.
RX MEDLINE=95011044; PubMed=7926278;
RA Gromoll J., Dankbar B., Gudermann T.;
RT "Characterization of the 5' flanking region of the human follicle-
RT stimulating hormone receptor gene."
RL Mol. Cell. Endocrinol. 102:93-102(1994).
RN [6]
3D-STRUCTURE MODELING OF 49-228.
RX MEDLINE=96363672; PubMed=8747461;
RA Jiang X., Drano M., Buckler D.R., Cheng S., Yehier A., Wu H.,
RA Hendrickson W.A., el Tayar N.;
RT "Structural predictions for the ligand-binding region of glycoprotein
RT hormone receptors and the nature of hormone-receptor interactions."
RL Structure 3:1341-1353(1995).
CC -!- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
CC OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC ADENYLATE CYCLASE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: A SHORT FORM OF THE TESTICULAR PROTEIN IS
CC PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE

```











Query Match 4.4%; Score 103; DB 1; Length 465;  
 Best Local Similarity 18.7%; Pred. No. 2.1;  
 Matches 76; Conservative 64; Mismatches 142; Indels 124; Gaps 16;

OY 110 GKILDSNNRINQYPSNAEYASLPDSLVKGVNWSAWALQCPKDA---SROYV 166  
 DB 79 GLILSVFQGVVQAP---VGYLSIYPMKRIIOSSLFSLMSLLIPPAQVGAALVI 135  
 OY 167 TCSNNIOAOQVIELARQNFET---PIDGLSSLSAREIENLPRLFTLWRGPVVV 218  
 DB 136 VCRVLQIAQAGTSGOHEIWKVWAPPLRGRLTS-----MTLSGYM--GPFIVLLVS 187  
 OY 219 -----AISLATEFFLYSFVRDVIHPYARNOQSDP----- 247  
 DB 188 GFICDLLGMPVFIYFGVGCVLVSLWFLFFDDPKD--HPTMSSSEKDYIISLMQOAS 245  
 OY 248 ---YKIPIEIVNKTPIVAITLLS----- 268  
 DB 246 SGROSLPKAMKSLPLWAILNSAFIWSNSLIVTYPTFTSTVJHNVNRENGLLSLP 305  
 OY 269 --LVYLAGLLAAAYOLYGYTKYRPPWLETWLOCRLGSLSPFFAMVHVAISCLPMR 326  
 DB 306 VLLAVICGILAGOMSEFELTRK-----IFSIVTVRKLTFTLGSFPCPVIFIMCLLYLSVN 359  
 OY 327 RSERVFLNMAVQOYHANIENSWNEEVRWREM---YISF--GIMSL-----GLLSLAV 376  
 DB 360 FYSTVIFUTLA-----NSTLSFYSGQLNALDIAPRYGFLKAATVALLGFMGGLLSITIA 415  
 OY 377 TSIPSVSNALNWRPESFTOSTGYVALISTFHLV-----IVGWRK 417  
 DB 416 GLILNODPEYAWKISFLMAGINVTCLV---FYFLFANGEIODWAK 458

## RESULT 11

RF3\_SACBA STANDARD; PRT; 476 AA.  
 AC P05512;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE MATURASE-LIKE RF3 PROTEIN.  
 GN RF3.  
 OS Saccharomyces bayanus (Yeast) (Saccharomyces uvarum).  
 OG Mitochondrion.  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
 OC Saccharomycetaceae; Saccharomycetes.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-87280035; PubMed-2440860;  
 RA Seraphin B., Simon M., Faye G.;  
 RT "The mitochondrial reading frame RF3 is a functional gene in  
 Saccharomyces uvarum."  
 RL J. Biol. Chem. 262:10146-10153(1987).  
 CC -!- SIMILARITY: THE RF3 PROTEINS ARE RELATED TO THE MATURASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; J03300; CAB23320.1; -;  
 DR PIR; A28439; A28439.  
 DR INTERPRO; IPR001982; -;  
 DR PFAM; PF00961; Intron\_maturase; 1.  
 KW Mitochondrion.  
 SQ SEQUENCE 476 AA; 57863 MW; ECB416C51DFFDA6D CRC64;

Query Match 4.3%; Score 100; DB 1; Length 476;  
 Best Local Similarity 20.7%; Pred. No. 3.6;

Matches 52; Conservative 48; Mismatches 85; Indels 66; Gaps 12;  
 OY 62 NPKFASFPFVVDVTHHEDALTGTNI-----IFVAIHREHYT-----SLWD 103  
 DB 254 NPYFVNAFSINI-----KTNLAKEIFTNLYNKLYSDYKINOINNHIPIYNYLK 302  
 OY 104 LHLVLGKILDSVNNMRINQYPSNAEYASLFF--PDSLVKGVNWSAWALQCPKDA 161  
 DB 303 INNKLPKINMDIKNNYWLAGTAAAGSFLSNYPKDTLLFKDM----- 347  
 OY 162 SROVVICSNIOAOQVIELARQNFETPIDGLSSLSAREIENLPRLFTLWRGPVVVAIS 221  
 DB 348 -RPSVIS-QVETRELILIOE---SFDL-SISNVAKVGNRKLKDFKLFRTTDELMK 400  
 OY 222 LATFFFLYSFVRDVIHPYARNOOSDFYKIP-----TEIVNKTLPVIAITLLSVYLAGLIA 277  
 DB 401 -----FIYF--DKFLPLHDKNQFVYKFERFTFKSYNNNRVFCVLSE--YNNIKI 451  
 OY 278 AYQLYYGTKY 288  
 DB 452 DNYDYVYNYKY 462

## RESULT 12

NU4M\_BRALA STANDARD; PRT; 452 AA.  
 AC O79421;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4 (EC 1.6.5.3).  
 GN ND4 OR NAD4 OR NADH4.  
 OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus).  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
 OC Branchiostoma.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-98292550; PubMed-9628930;  
 RA Spruyt N., Delarbre C., Gachelin G., Laudet V.;  
 RT "Complete sequence of the amphioxus (Branchiostoma lanceolatum)  
 mitochondrial genome: relations to vertebrates."  
 RL Nucleic Acids Res. 26:3279-3285(1998).  
 CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Y16474; CAA76256.1; -;  
 DR INTERPRO; IPR000260; -;  
 DR INTERPRO; IPR001750; -;  
 DR PFAM; PF00361; oxidored\_g1; 1.  
 DR PFAM; PF01059; oxidored\_g5; 1.  
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.  
 SQ SEQUENCE 452 AA; 49940 MW; 4067DB59EC184C7A CRC64;

Query Match 4.2%; Score 99.5; DB 1; Length 452;  
 Best Local Similarity 20.1%; Pred. No. 3.7;  
 Matches 95; Conservative 56; Mismatches 139; Indels 183; Gaps 25;  
 OY 109 VQKIL-----IDVSNMNRINQYPSNAEYASLFFPDSLVKGVNWSAWALQCPKAS 162  
 DB 29 VGSVLLMLPATVLVNNMTI-----SNVSMMYT---SDFVSLGLTVLSITWLLPL-MLLAS 79  
 OY 163 ROVYICSNIOAR-----QOVIELARQNFETIDGLSSLSAREIENLP-IRLFTLWRGPV 216  
 DB 80 OOHMVSESLIYQRFVGCQVFLTGALVLAFAFMSDILLFYIAFESTLLPTLMLTRW-GAQ 138





DR PIR: S07757; GNVVYP.  
 DR HSP: P14336; LSVB.  
 DR MEROPS; S07.001; -  
 DR INTERPRO: IPR000069; -  
 DR INTERPRO: IPR000208; -  
 DR INTERPRO: IPR000336; -  
 DR INTERPRO: IPR000404; -  
 DR INTERPRO: IPR000487; -  
 DR INTERPRO: IPR000752; -  
 DR INTERPRO: IPR001122; -  
 DR INTERPRO: IPR001157; -  
 DR INTERPRO: IPR001528; -  
 DR INTERPRO: IPR001850; -  
 DR INTERPRO: IPR002535; -  
 DR PFAM: PF01004; Flavi\_LM; 1.  
 DR PFAM: PF00948; Flavi\_NSL; 1.  
 DR PFAM: PF01005; Flavi\_NS2A; 1.  
 DR PFAM: PF01002; Flavi\_NS2B; 1.  
 DR PFAM: PF01350; Flavi\_NS4A; 1.  
 DR PFAM: PF01349; Flavi\_NS4B; 1.  
 DR PFAM: PF00972; Flavi\_NS5; 1.  
 DR PFAM: PF01003; Flavi\_capsid; 1.  
 DR PFAM: PF00869; Flavi\_glycoprot; 1.  
 DR PFAM: PF01570; Flavi\_propep; 1.  
 DR PFAM: PF00949; Flavi\_helicase; 1.  
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein.  
 FT INIT\_MET 1 1  
 FT CHAIN 1 121  
 FT PROPEP 122 210  
 FT CHAIN 211 285  
 FT CHAIN 286 778  
 FT CHAIN 779 1130  
 FT CHAIN 1131 1354  
 FT CHAIN 1355 1484  
 FT CHAIN 1485 2107  
 FT CHAIN 2108 2256  
 FT CHAIN 2257 2506  
 FT CHAIN 2507 3411  
 FT NP\_BIND 1682 1689  
 FT SITE 1773 1776  
 FT TRANSMEM 249 269  
 FT TRANSMEM 271 285  
 FT TRANSMEM 740 753  
 FT TRANSMEM 755 778  
 FT TRANSMEM 1159 1180  
 FT DISULFID 288 315  
 FT DISULFID 345 401  
 FT DISULFID 359 390  
 FT DISULFID 377 406  
 FT DISULFID 467 568  
 FT DISULFID 585 615  
 FT CARBOHYD 134 134  
 FT CARBOHYD 150 150  
 FT CARBOHYD 908 908  
 FT CARBOHYD 986 986  
 FT CARBOHYD 2320 2320  
 FT CARBOHYD 2346 2346  
 FT CARBOHYD 2467 2467  
 SQ SEQUENCE 3411 AA; 379524 MW; 3298C0771FED23F7 CRC64;

Query Match 4.28; Score 98; DB 1; Length 3411;  
 Best Local Similarity 22.2%; Pred. No. 49;  
 Matches 78; Conservative 55; Mismatches 130; Indels 88; Gaps 20;  
 QY 107 LLVGKI-LID-----VSNNNRINQYPE-SNAEYLA-----SLFPDSLVKGFNVVSAWA-153  
 Db 1171 MLVGQVTLDLKLTAVAGLHFHFMNNGGDAMKALIAFPIRPLLI--GFGLRLWSP 1228  
 QY 154 -----LQLGPKDASRQVYICSNINQARQQVIELARQLNFTIPIDLGSLS--SAREIENLPL 206

Db 1229 RERLVLTG-----AAMVEIALGGVMG-----GLWKVNLNAVSLCILINAVASRKASNTIL 1279  
 QY 207 RLFTLWRGPVVVAISLATFEFLYSFYVDVHPYARNOQSDFYKIPTEIVNKTLPVIAITL 266  
 Db 1280 PLMALITPVTMAEVRLAAMFEFCAVVIIGVLRHQFKDTS-----MOKTIPLVALT 1329  
 QY 267 LSLVYLA---GLLAAAYQLYGTGYRRPFPWLETWLCQRKQLGLLSFFPFAMVHVAYSLC 322  
 Db 1330 TSYLGITQPELGLCAFLATRIFG---RRSIPVNEA-LAAAGLVGVLA-----G 1373  
 QY 323 LPMRRSERYL-----FLNWAYQQVHANI-----ENSWREE-EVNRIMYISFGIM 366  
 Db 1374 LAFQEMENELCPITAVGGLLMLVSVAGKVDGLLELKKLGEVSWEEAEISGSSARYDVALS 1433  
 QY 367 SLGLLSLLAVTSIPSVSNALNWREFSEFIQSTLCYVALLISTPHVLIY--GW 415  
 Db 1434 EQGEFKLLSEKVP-----WDO--VVMTSLALVGAALHPFALLLVLAGM 1475

Search completed: March 14, 2001, 16:15:16  
 Job time: 148 sec